(D2)

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) Publication number: **0 661 688 A2**

(12) EUROPEAN PATENT APPLICATION

(21) Application number: **94308600.9**

(51) Int. Cl.[6]: **G10L 3/00**, G10L 5/06

(22) Date of filing: **22.11.94**

(30) Priority: **30.12.93 US 175701**

(43) Date of publication of application:
05.07.95 Bulletin 95/27

(84) Designated Contracting States:
DE FR GB

(71) Applicant: **International Business Machines Corporation**
Old Orchard Road
Armonk, N.Y. 10504 (US)

(72) Inventor: **Cohen, Paul S.**
3271 Nutly Circle
Yorktown Heights,
New York 10598 (US)
Inventor: **Lucassen, John M.**
308 West 103rd Street
New York,
New York 10025 (US)
Inventor: **Miller, Roger M.**
P O Box 778
Brookfield,
Connecticut 06804 (US)
Inventor: **Sherwin, Elton B.**
26 Dogwood Lane
Stamford,
Connecticut 06903 (US)

(74) Representative: **Burt, Roger James, Dr.**
IBM United Kingdom Limited
Intellectual Property Department
Hursley Park
Winchester
Hampshire SO21 2JN (GB)

(54) System and method for location specific speech recognition.

(57) A system and method are disclosed for reducing perplexity in a speech recognition system based upon determined geographic location. In a mobile speech recognition system which processes input frames of speech against stored templates representing speech, a core library of speech templates is created and stored representing a basic vocabulary of speech. Multiple location-specific libraries of speech templates are also created and stored, each library containing speech templates representing a specialized vocabulary for a specific geographic location. The geographic location of the mobile speech recognition system is then periodically determined utilizing a cellular telephone system, a geopositioning satellite system or other similar systems, and a particular one of the location-specific libraries of speech templates is identified for the current location of the system. Input frames of speech are then processed against the combination of the core library and the particular location-specific library to greatly enhance the accuracy and efficiency of speech recognition by the system. Each location-specific library preferably includes speech templates representative of location place names, proper names, and business establishments within a specific geographic location.

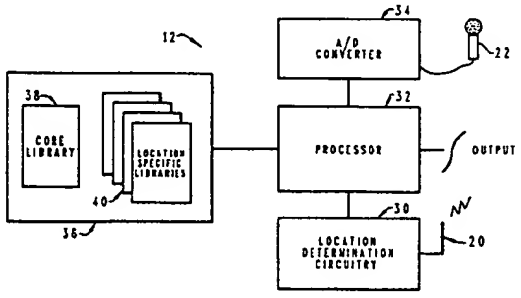

Fig. 2

EP 0 661 688 A2

The present invention relates in general to speech recognition systems and in particular to a system and method for enhancing speech recognition accuracy in a mobile speech recognition system.

Speech recognition is well known in the prior art. The recognition of isolated words from a given vocabulary for a known speaker is perhaps the simplest type of speech recognition and this type of speech recognition has been known for some time. Words within the vocabulary to be recognized are typically prestored as individual templates, each template representing the sound pattern for a word in the vocabulary. When an isolated word is spoken, the system merely compares the word to each individual template which represents the vocabulary. This technique is commonly referred to as whole-word template matching. Many successful speech recognition systems use this technique with dynamic programming to cope with nonlinear time scale variations between the spoken word and the prestored template.

Of greater difficulty is the recognition of continuous speech or speech which contains proper names or place names. Continuous speech, or connected words, have been recognized in the prior art utilizing multiple path dynamic programming. One example of such a system is proposed in "Two Level DP Matching A Dynamic Programming Based Pattern Matching Algorithm For Connected Word Recognition" H. Sakoe, IEEE Transactions on Acoustics Speech and Signal Processing, Volume ASSP-27, No. 6, pages 588-595, December 1979. This paper suggests a two-pass dynamic programming algorithm to find a sequence of word templates which best matches the whole input pattern. Each pass through the system generates a score which indicates the similarity between every template matched against every possible portion of the input pattern. In a second pass the score is then utilized to find the best sequence of templates corresponding to the whole input pattern.

United States Patent No. 5,040,127 proposes a continuous speech recognition system which processes continuous speech by comparing input frames against prestored templates which represent speech and then creating links between records in a linked network for each template under consideration as a potentially recognized individual word. The linked records include ancestor and descendent link records which are stored as indexed data sets with each data set including a symbol representing a template, a sequence indicator representing the relative time the link record was stored and a pointer indicating a link record in the network from which it descends.

The recognition of proper names represents an increase in so-called "perplexity" for speech recognition systems and this difficulty has been recently recognized in U.S. Patent No. 5,212,730. This patent performs name recognition utilizing text-derived recognition models for recognizing the spoken rendition of proper names which are susceptible to multiple pronunciations. A name recognition technique set forth within this patent involves entering the name-text into a text database which is accessed by designating the name-text and thereafter constructing a selected number of text-derived recognition models from the name-text wherein each text-derived recognition model represents at least one pronunciation of the name. Thereafter, for each attempted access to the text database by a spoken name input the text database is compared with the spoken name input to determine if a match may be accomplished.

U.S. Patent No. 5,202,952 discloses a large-vocabulary continuous-speech prefiltering and processing system which recognizes speech by converting the utterances to frame data sets wherein each frame data set is smoothed to generate a smooth frame model over a predetermined number of frames. Clusters of word models which are acoustically similar over a succession of frame periods are designated as a resident vocabulary and a cluster score is then generated by the system which includes the likelihood of the smooth frames evaluated utilizing a probability model for the cluster against which the smoothed frame model is being compared.

Each of these systems recognizes that successful speech recognition requires a reduction in the perplexity of a continuous-speech utterance. Publications which address this problem are "Perplexity-A Measure of Difficulty of Speech Recognition Tasks," Journal of the Acoustical Society of America, Volume 62, Supplement No. 1, page S-63, Fall 1977, and the "Continuous Speech Recognition Statistical Methods" in the Handbook of Statistics Volume 2: Classification, Pattern Recognition and Reduction of Dimensionality, pages 549-573, North-Holland Publishing Company, 1982.

In view of the above, it is apparent that successful speech recognition requires an enhanced ability to distinguish between large numbers of like sounding words, a problem which is particularly difficult with proper names, place names and numbers.

It is therefore an object of the present invention to provide a speech recognition system, and method of operation of such a system, with enhanced speech recognition accuracy and efficiency.

Accordingly the present invention provides a mobile speech recognition system comprising: an audio input means for receiving input speech; a

storage means; a core library of speech templates stored within said storage means representing a basic vocabulary of speech; a plurality of location-specific libraries of speech templates stored within said storage means, each representing a specialized vocabulary for a particular geographic location; location determination means for determining a geographic location of said mobile speech recognition system; library selection means coupled to said storage means and said location determination means for selecting a particular one of said plurality of location-specific libraries in response to a determination of said geographic location of said mobile speech recognition system; and speech processor means coupled to said audio input means and said storage means for processing input speech frames against said core library and said particular one of said plurality of location-specific libraries.

Viewed from a second aspect the present invention provides a method of operating a mobile speech recognition system to process input frames of speech against stored templates representing speech, said method comprising the steps of: storing in a memory a core library of speech templates representing a basic vocabulary of speech; storing a plurality of location-specific libraries of speech templates, each representing a specialized vocabulary for a particular geographic location; determining a geographic location of said mobile speech recognition system; associating said core library of speech templates with a particular one of said plurality of location-specific libraries of speech templates in response to said determination of said geographic location of said mobile speech recognition system; and employing a processor to process input frames of speech against said core library and associated location-specific library.

From the above it is apparent that the present invention provides a system and method for enhanced speech recognition in a mobile system utilizing location-specific libraries of speech templates and an identification of the system location.

The system and method of the preferred embodiment of the invention reduce perplexity in a speech recognition system based upon determined geographic location. In a mobile speech recognition system according to the preferred embodiment input frames of speech are processed against stored templates representing speech, a core library of speech templates being created and stored representing a basic vocabulary of speech. Multiple location-specific libraries of speech templates are also created and stored, each library containing speech templates representing a specialized vocabulary for a specific geographic location. The geographic location of the mobile speech recognition system is then periodically determined utilizing a cellular telephone system, a geopositioning satellite system or other similar systems and a particular one of the location-specific libraries of speech templates is identified for the current location of the system. Input frames of speech are then processed against the combination of the core library and the particular location-specific library to greatly enhance the efficiency of speech recognition by the system. Each location-specific library preferably includes speech templates representative of location place names, proper names, and business establishments within a specific geographic location.

The present invention will be described further, by way of example only, with reference to a preferred embodiment thereof as illustrated in the accompanying drawings, in which:

Figure 1 is a pictorial representation of a mobile speech recognition system which may be utilized to implement the system and method of the present invention;

Figure 2 is a high-level block diagram of the mobile speech recognition system of **Figure 1**; and

Figure 3 is a high-level logic flowchart illustrating a process for implementing the method of the present invention.

With reference now to the figures and in particular with reference to **Figure 1**, there is depicted a pictorial representation of a mobile speech recognition system **12** which may be utilized to implement the system and method of the preferred embodiment of the present invention. As illustrated, mobile speech recognition system **12** may be implemented utilizing any suitably programmed portable computer, such as a so-called "notebook computer." As depicted, mobile speech recognition system **12** may include a keyboard **14**, a display **16** and a display screen **18**. Additionally, as will be explained in greater detail herein, mobile speech recognition system **12** may also include an antenna **20** which may be utilized to electronically determine the specific geographic location of mobile speech recognition system **12** in response to detection of a verbal utterance.

Also depicted within **Figure 1** is an audio input device which is coupled to mobile speech recognition system **12**. Microphone **22** serves as an audio input device for mobile speech recognition system **12** and, in a manner well known to those having ordinary skill in the speech recognition art, may be utilized to capture verbal utterances spoken by a user of mobile speech recognition system **12**, in order to provide additional information, perform specific functions or otherwise respond to verbal commands.

Mobile speech recognition system **12** is characterized as mobile within this specification and it

is anticipated that such systems will find application within mobile platforms, such as automobiles, police cars, fire trucks, ambulances, and personal digital assistant (PDAs) which may be carried on the person of a user. Upon reference to this disclosure, those skilled in the art will appreciate that speech recognition on a mobile platform represents a definite increase in the likely perplexity of the speech recognition problem due to the necessity that the system recognize street names, street addresses, restaurant names, business names and other proper names associated with a specific geographic location at which the mobile system may be located.

In order to solve this problem mobile speech recognition system **12** preferably includes a device for determining the geographic location of the system. This may be accomplished utilizing many different techniques including the utilization of a geopositioning system, such as the Global Positioning Satellite System. Thus, radio signals from satellite **28** may be received by mobile speech recognition system **12** at antenna **20** and may be utilized to determine the specific geographic location for mobile speech recognition system **12** at the time of a particular spoken utterance. Similarly, radio signals from a cellular phone network, such as cellular transmission towers **24** and **26**, may also be utilized to accurately and efficiently determine the geographic location of mobile speech recognition system **12**. Additionally, while not illustrated, those skilled in the art will appreciate that special purpose radio signals, inertial guidance systems, or other similar electronic measures may be utilized to determine the geographic location of mobile speech recognition system **12** in a manner that is well within the current scope of these technologies. Additionally, a user may simply enter an indication of geographic location into mobile speech recognition system **12** utilizing keyboard **14**.

Referring now to **Figure 2**, there is depicted a high level block diagram of the mobile speech recognition system **12** of **Figure 1**, which illustrates the manner in which this geographic location determination may be utilized to decrease the perplexity of speech recognition. As illustrated within **Figure 2**, a memory **36** is provided within mobile speech recognition system **12** which includes a core library **38** of speech templates which represent a basic vocabulary of speech. Similarly, multiple location-specific libraries **40** are also stored within memory **36**. Each location-specific library **40** includes templates which are representative of a specialized vocabulary for a particular geographic location. For example, each location-specific library of speech templates may include a series of speech templates representative of street names within that geographic location, business establishments within that geographic location or other proper names which are germane to a selected geographic location.

Thus, each time a speech utterance is detected at microphone **22**, that utterance may be suitably converted for processing utilizing analog-to-digital converter **34** and coupled to processor **32**. Processor **32** then utilizes location determination circuitry **30** to identify a specific geographic location for mobile speech recognition system **12** at the time of the spoken utterance. This may be accomplished, as described above, utilizing global positioning satellite systems or radio frequency signals from cellular telephone systems which are received at antenna **20**, or by the simple expedient of requiring the user to periodically enter at keyboard **14** an identification for a specific geographic location.

Next, the output of location determination circuitry **30** is utilized by processor **32** to select a particular one of the multiple location-specific libraries **40** contained within memory **36**. The input frame of speech data is then compared to a composite library which is comprised of core library **38** and a particular one of the location-specific libraries **40**. In this manner, the perplexity of speech recognition in a mobile speech recognition system may be greatly reduced, thereby enhancing the accuracy and efficiency of speech recognition within the system.

As discussed above with respect to previous attempts at speech recognition, the templates against which input speech is processed may comprise templates representing individual words, phrases or portions of words. As utilized herein, the term "template" shall mean any stored digital representation which may be utilized by processor **32** to identify an unknown speech utterance.

Finally, with reference to **Figure 3**, there is depicted a high-level logic flowchart which illustrates a process for implementing the method of the preferred embodiment of the present invention. As depicted, this process begins at block **60** and thereafter passes to block **62**. Block **62** illustrates a determination of whether or not a verbal utterance has been detected. If not, the process merely iterates until such time as an utterance has been detected. However, once a verbal utterance has been detected, the process passes to block **64**.

Block **64** illustrates a determination of the current geographic location of mobile speech recognition system **12**. As discussed above, this may be accomplished utilizing the global positioning satellite system, cellular telephone systems, or other specialized radio frequency signals or inertial navigation techniques. Thereafter, the geographic location determination is utilized to select a particular

location-specific library, as depicted at block **66**.

Next, as depicted at block **68**, the input utterance is processed against the core library of basic vocabulary words and a particular location-specific library which is associated with the determined geographic location of mobile speech recognition system **12**. Thereafter, the process passes to block **70**. Block **70** illustrates a determination of whether or not the verbal utterance has been recognized. If not, the process passes to block **72** which depicts the generation of an error message, a verbalized command urging the user to repeat the utterance, or other similar techniques for resolving the failure of the system to recognize the verbal utterance. After generating such a message, the process then passes to block **76** and returns to await the detection of a subsequent verbal utterance.

Referring again to block **70**, in the event the utterance has been recognized, the process passes to block **74**. Block **74** illustrates the processing of that utterance. Those skilled in the art will appreciate that a verbal utterance may be processed to generate information which is presented to the user, to control the activity of some portion of the system or to store data for future reference. Thereafter, as above, the process passes to block **76** and returns to await the detection of a subsequent verbal utterance.

Upon reference to the foregoing, those skilled in the art will appreciate that by determining the geographic location of a mobile speech recognition system and thereafter utilizing a location-specific library of speech templates, the method and system of the preferred embodiment of the present invention greatly reduces the possible perplexity of a speech recognition system and concomitantly enhances the accuracy and efficiency of speech recognition.

Claims

1. A mobile speech recognition system comprising:

   an audio input means (22) for receiving input speech;

   a storage means (36);

   a core library (38) of speech templates stored within said storage means (36) representing a basic vocabulary of speech;

   a plurality of location-specific libraries (40) of speech templates stored within said storage means (36), each representing a specialized vocabulary for a particular geographic location;

   location determination means (30) for determining a geographic location of said mobile speech recognition system;

   library selection means (32) coupled to said storage means (36) and said location determination means (30) for selecting a particular one of said plurality of location-specific libraries in response to a determination of said geographic location of said mobile speech recognition system; and

   speech processor means (32) coupled to said audio input means (22) and said storage means (36) for processing input speech frames against said core library and said particular one of said plurality of location-specific libraries.

2. A mobile speech recognition system as claimed in claim 1, wherein said audio input means (22) comprises a microphone.

3. A mobile speech recognition system as claimed in claim 1 or claim 2, wherein said location determination means (30) comprises a cellular telephone transceiver.

4. A mobile speech recognition system as claimed in claim 1 or claim 2, wherein said location determination means (30) comprises a global positioning satellite receiver.

5. A mobile speech recognition system as claimed in any preceding claim, wherein said speech processor means (32) includes an analog-to-digital converter (34).

6. A mobile speech recognition system as claimed in any preceding claim, wherein said speech processor means (32) forms part of a personal computer.

7. A mobile speech recognition system as claimed in any preceding claim, wherein each of said plurality of location-specific libraries of speech templates comprises a plurality of templates representative of a plurality of location place names.

8. A method of operating a mobile speech recognition system to process input frames of speech against stored templates representing speech, said method comprising the steps of:

   storing in a memory (36) a core library of speech templates representing a basic vocabulary of speech;

   storing a plurality of location-specific libraries of speech templates, each representing a specialized vocabulary for a particular geographic location;

   determining (64) a geographic location of said mobile speech recognition system;

   associating (66) said core library of speech templates with a particular one of said plurality

of location-specific libraries of speech templates in response to said determination of said geographic location of said mobile speech recognition system; and

employing a processor (32) to process (68) input frames of speech against said core library and associated location-specific library.

9. A method as claimed claim 8, wherein said step of determining (64) a geographic location of said mobile speech recognition system comprises the step of utilizing a cellular telephone system to determine a geographic location of said mobile speech recognition system.

10. A method as claimed in claim 8, wherein said step of determining a geographic location of said mobile speech recognition system comprises the step of utilizing a geopositioning satellite system receiver to determine a geographic location of said mobile speech recognition system.

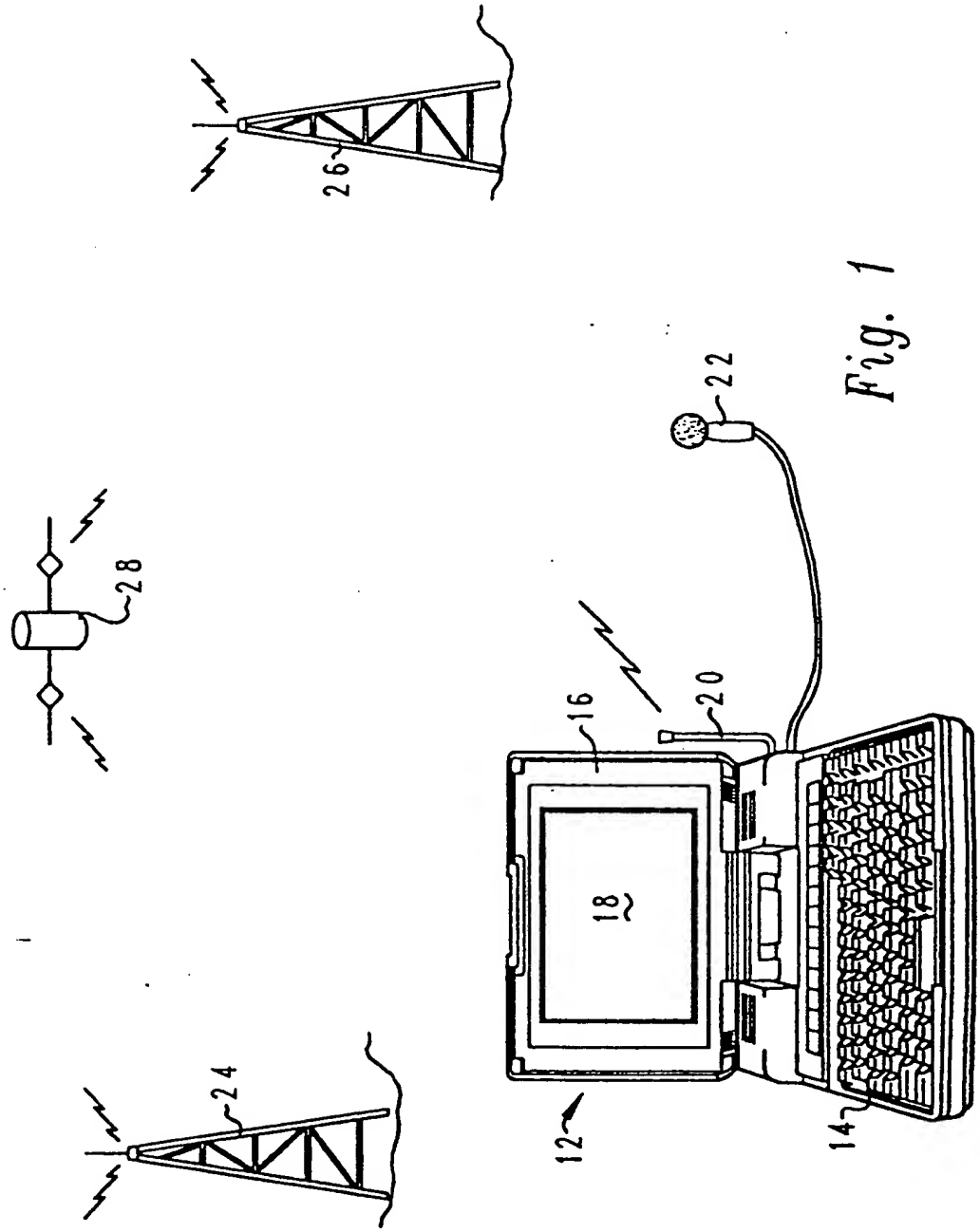

Fig. 1

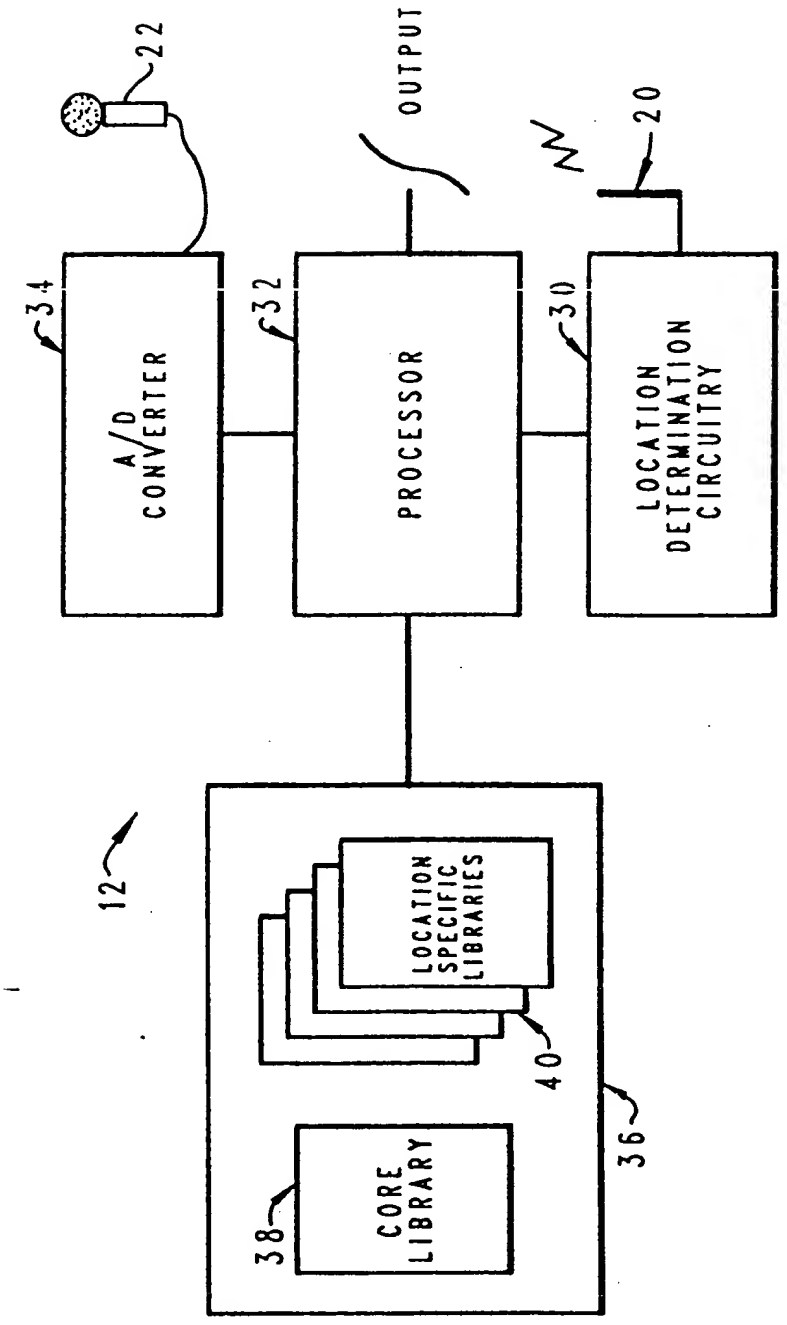
22
OUTPUT
20
34
32
30
A/D CONVERTER
PROCESSOR
LOCATION DETERMINATION CIRCUITRY
12
LOCATION SPECIFIC LIBRARIES
40
36
38
CORE LIBRARY

Fig. 2

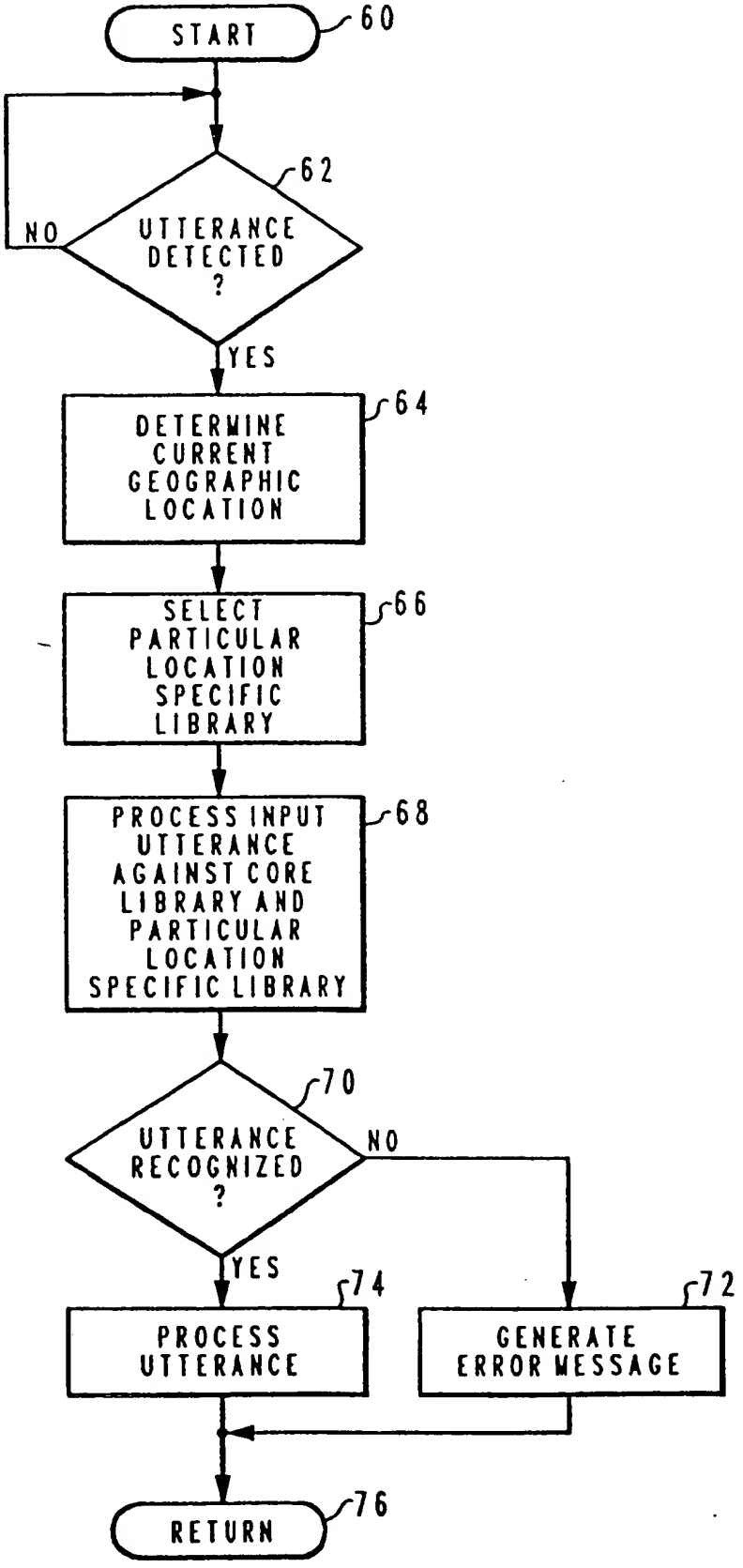
START
60
62
UTTERANCE DETECTED ?
NO
YES
DETERMINE CURRENT GEOGRAPHIC LOCATION
64
SELECT PARTICULAR LOCATION SPECIFIC LIBRARY
66
PROCESS INPUT UTTERANCE AGAINST CORE LIBRARY AND PARTICULAR LOCATION SPECIFIC LIBRARY
68
70
UTTERANCE RECOGNIZED ?
NO
YES
74
PROCESS UTTERANCE
72
GENERATE ERROR MESSAGE
RETURN
76

Fig. 3